Volume 13 • Number 4 • 2010
VALUE IN HEALTH

Bayesian Analysis of Nosocomial Infection Risk and Length of Stay in a Department of General and Digestive Surgery

Antonio José Sáez-Castillo, PhD,[1] María José Olmo-Jiménez, PhD,[1] José María Pérez Sánchez, PhD,[2] Miguel Ángel Negrín Hernández, PhD,[3] Ángel Arcos-Navarro, PhD,[4] Juan Díaz-Oller, PhD[4]

[1]Department of Statistics and Operational Research, University of Jaén, Spain; [2]Department of Quantitative Methods for Economics and Business, University of Granada, Spain; [3]Department of Quantitative Methods in Economics and Management, University of Las Palmas de Gran Canaria, Spain; [4]Department of General and Digestive Surgery, San Agustín Hospital, Linares, Spain

ABSTRACT

Objective: Nosocomial infection is one of the main causes of morbidity and mortality in patients admitted to hospital. One aim of this study is to determine its intrinsic and extrinsic risk factors. Nosocomial infection also increases the duration of hospital stay. We quantify, in relative terms, the increased duration of the hospital stay when a patient has the infection.
Methods: We propose the use of logistic regression models with an asymmetric link to estimate the probability of a patient suffering a nosocomial infection. We use Poisson-Gamma regression models as a multivariate technique to detect the factors that really influence the average hospital stay of infected and noninfected patients. For both models, frequentist and Bayesian estimations were carried out and compared.

Results: The models are applied to data from 1039 patients operated on in a Spanish hospital. Length of stay, the existance of a preoperative stay and obesity were found the main risk factors for a nosomial infection. The existence of a nosocomial infection multiplies the length of stay in the hospital by a factor of 2.87.
Conclusion: The results show that the asymmetric logit improves the predictive capacity of conventional logistic regressions
Keywords: asymmetric logit, Bayesian analysis, length of stay in hospital, logistic regression, nosocomial infection risk, Poisson-Gamma model.

1. Introduction

Nosocomial infections (NI) are infections that develop during hospitalization and are neither present nor incubating at the time of the patient's admission. Currently, hospital infection or NI remains a major problem, constituting one of the main causes of morbidity and mortality in patients admitted to hospital. Although the figure varies considerably among countries, some studies estimate that approximately one in ten hospitalized patients will acquire an infection after admission [1]. In Spain, the overall prevalence rate of patients with NI has decreased from 8.5% in 1990 to 7% in 2007 [2–4].

For this reason, determining the intrinsic and extrinsic risk factors to which these patients are exposed and predicting NI are important aims of research. Furthermore, NI clearly increases the duration of hospital stay, causing direct economic costs and other costs derived from specific laboratory and isolation techniques and from lengthy antibiotic treatments. Estimates of the cost of these infections, in 2002 prices, suggest that the annual economic burden is $6.7 billion per year in the United States [5] and £1.06 billion in the United Kingdom [6].

In view of the foregoing, the first aim of this article is to estimate the risk factors for NI in a hospital's general surgery and digestive department ([7–10], among others). One of the statistical techniques that has traditionally been used to predict NI is the logistic regression, which not only allows the effect of each risk factor to be evaluated, but also makes it possible to quantify the NI probability of a given patient. We carried out the Bayesian estimation of these regression models. Recently, there has been great interest in Bayesian regression techniques for dichotomous response variables in many fields of application [11–16]. Chen et al. [17] also apply a Bayesian approach in their proposal to use an asymmetric link for analyzing binary response data when one response is much more frequent than the other. We compare the results of applying a Bayesian estimation with those obtained by the frequentist estimation for logistic regression models.

Patients with hospital-acquired infections suffer a prolonged stay, during which time they occupy scarce bed-days and require additional diagnostic and therapeutic interventions [18]. As a second objective of this study, we set out to determine the factors that influence hospital stay, using a Poisson-Gamma regression model. A particular aim is to quantify, in relative terms, the increased duration of the hospital stay when a patient has NI. Frequentist and Bayesian estimations for this model are compared.

The article is organized as follows: section 2 describes the data, introducing the covariates used in the study and section 3 addresses the analysis of the methodology to be considered. The results of the article are shown in section 4, and section 5 is devoted to a discussion of the results and to summarizing the conclusions reached.

2. Data

Data were collected in a prospective cohort study of 1039 patients operated on between January 1, 1998 and December 31, 1998 at the General and Digestive Surgery Department of the North Area Hospital in the province of Jaén (Spain). Only patients of first admission and with at least 1 day of hospitalization were considered.

NI was defined as any infection that was active or under antibiotic treatment and that occurred 48 hours after the hospitalization [19]. Patients were followed up for 1 month after hospital discharge.

Address correspondence to: Miguel Ángel Negrín-Hernández, Department of Quantitative Methods in Economics and Management, Faculty of Economics and Business, Campus de Tafira, 35017, Las Palmas de Gran Canaria, Spain. E-mail: mnegrin@dmc.ulpgc.es
10.1111/j.1524-4733.2009.00680.x

Table 1 Descriptive summary of quantitative variables

Variable	Min	Max	Mean	P_{25}	P_{50}	P_{75}
Length of stay	1	73	5.27	1	2	5
Age	7	88	49.56	35	50	66
Length of surgery	3	460	65.76	30	50	80
Preoperative stay	0	34	0.99	0	0	0
Number of diagnoses	1	6	1.68	1	1	2

We consider both intrinsic and extrinsic risk factors for NI. The intrinsic factors are patient related and the extrinsic factors are related to medical intervention. The intrinsic factors considered were age, sex (male = 1 and female = 0), the presence or absence in each patient of coma, kidney failure, diabetes, neoplasy, chronic obstructive pulmonary disease, chronic hepatopathy, immunodeficiency, hypoproteinemia, obesity, and infection at admission, which includes NI due to a previous admission in the same hospital.

During the patients' hospital stay, the type of admission (scheduled = 1 and urgent = 0) and the presence or absence of the following extrinsic factors was recorded: peripheral tract, central tract, vesical probe, nasogastric probe, open drainage, closed drainage, artificial respiration, and immunosuppressive therapy. With regard to diagnosis-related data, the total number of diagnoses, based on important diagnosis, no symptoms, or isolated signs, was considered. Finally, the following factors related to surgery were taken into account: surgery type (scheduled = 1 or urgent = 0), length of surgery (in minutes), existence of antibiotic prophylaxis, preoperative stay, and degree of contamination, with four categories (always related to the main surgery method if several were applied): clean, clean-contaminated, contaminated, and dirty surgery. The total hospital stay (in days) was also recorded. A descriptive study of all these variables in the sample is shown in Tables 1–3.

3. Methodology

Firstly, we propose two alternative discrete choice models to predict the probability of NI. Symmetric and asymmetric links are considered, together with frequentist and Bayesian approaches. Secondly, Poisson-Gamma regression models are proposed to estimate the extension of the hospital stay caused by NI.

Table 2 Descriptive summary of categorical variables (absence or presence)

Variable	Yes (1)	No (0)
NI	64 (6.27%)	957 (93.73%)
Prophylaxis	803 (78.65%)	218 (21.35%)
Peripheral tract	1019 (99.80%)	2 (0.20%)
Central tract	82 (8.03%)	939 (91.97%)
Vesical probe	191 (18.71%)	830 (81.29%)
Nasogastric probe	185 (18.12%)	836 (81.88%)
Open drainage	397 (38.88%)	624 (61.12%)
Closed drainage	118 (11.56%)	903 (88.44%)
Artificial respiration	19 (1.86%)	1002 (98.14%)
Immunosuppressive therapy	19 (1.86%)	1002 (98.14%)
Coma	18 (1.76%)	1003 (98.24%)
Kidney failure	10 (0.98%)	1011 (99.02%)
Diabetes	104 (10.19%)	917 (89.81%)
Neoplasy	91 (8.91%)	930 (91.09%)
COPD	111 (10.87%)	910 (89.13%)
Chronic hepatopathy	39 (3.82%)	982 (96.18%)
Immunodeficiency	7 (0.69%)	1014 (99.31%)
Hypoproteinemia	29 (2.84%)	992 (97.16%)
Infection in admission	177 (17.34%)	844 (82.66%)
Obesity	148 (14.50%)	873 (85.50%)

COPD, chronic obstructive pulmonary disease; NI, nosocomial infection.

NI Predictive Models

Let $\mathrm{y} = (y_1, y_2, \ldots, y_n)'$ denote an $\mathrm{n} \times 1$ vector of a dependent dichotomic variable and $\mathrm{x}_i = (x_{i1}, \ldots, x_{ik})'$ denote the $k \times 1$ vector of covariates for the patient i. A predictive regression model deals with the problem of estimating the binary variable y_i, which represents the fact of belonging or not to a study group. In this case, $y_i = 1$ if the ith individual suffers an NI, and $y_i = 0$ otherwise. Assume that $y_i = 1$ with probability p_i and $y_i = 0$ with probability $1 - p_i$. In this dichotomous model, x_i includes the risk factors for the ith individual. The regression model is given by

$$p_i = F(\mathrm{x}_i'\beta)$$

where $\beta = (\beta_1, \ldots, \beta_K)'$ is a $k \times 1$ vector of regression coefficients, which represents the effect of each factor in the model and $F(\cdot)$ is the link function. The likelihood function is given by

$$l(\mathrm{y}|\mathrm{x}, \beta) = \prod_{i=1}^{n} [F(\mathrm{x}_i'\beta)]^{y_i} [1 - F(\mathrm{x}_i'\beta)]^{1-y_i} \quad (1)$$

where $\mathrm{x} = (\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n)'$.

Frequentist estimation of conventional logit models. For conventional logistic regression, the link function is equal to $F(z) = \dfrac{1}{1+e^{-z}}$. Observe that this is a symmetric function with respect to zero, so $F(-z) = 1 - F(z)$ for all z.

The regression coefficients, β, are usually estimated by numerical evaluation of the likelihood function. Then, the model provides the probability of infection for any individual. The normal procedure is then to consider a cutoff in this probability for detecting infected individuals.

Bayesian estimation of symmetric and asymmetric logit models. A Bayesian estimation of the logistic regression model is obtained by assuming that the β coefficients are random nodes of the model. To facilitate the comparison with frequentist methods of estimation, we assume centered and noninformative normal densities as prior distributions for the coefficients.

We also propose the use of an asymmetric link function, fitting the resulting model from a Bayesian point of view. The model has been used in other contexts ([16,17,20,21], among others), but has had little application in the health field. The asymmetric model is adequate for binary response data when one response is much more frequent than the other, as occurs in the case we examine in this study.

Following Albert and Chib [11] and Chen et al. [17], we assume that the model uses a vector of latent variables $\mathrm{w} = (w_1, w_2, \ldots, w_n)'$ in this form:

$$y_i = \begin{cases} 0, & w_i \le 0, \\ 1, & w_i \ge 0, \end{cases}$$

where

$$w_i = x_i'\beta + \delta z_i + \varepsilon_i, \quad z_i \sim G, \varepsilon_i \sim F$$

In this model, G is the cumulative distribution function of the half-standard normal distribution given by

$$g(z) = \frac{2}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, z > 0$$

Table 3 Descriptive summary of categorical variables

Variable	Categories			
Sex	Male 607 (59.45%)	Female* 414 (40.55%)		
Admission	Scheduled 631 (61.80%)	Urgent* 390 (38.20%)		
Surgery type	Scheduled 671 (65.72%)	Urgent* 350 (34.28%)		
Degree of contamination	Clean* 437 (42.80%)	Clean-contaminated 164 (16.06%)	Contaminated 111 (10.87%)	Dirty 309 (30.26%)

*Indicates the reference category.

F is the standard logistic cumulative distribution function, and z_i and ε_i are assumed to be independent. The skewness in this regression model is given by δz_i, where $\delta \in (-\infty, \infty)$ is the skewness parameter. If $\delta < 0$ then the probability of $p_i = 0$ increases, although if $\delta > 0$, the probability of $p_i = 1$, i.e., the infection probability of the ith individual, increases. Obviously, if $\delta = 0$, then the regression model is reduced to a standard logit.

The likelihood function in Eq. 1 can be rewritten as

$$l(y|\mathrm{x}, \beta, \delta) = \prod_{i=1}^{n} \int_0^{\infty} [F(\mathrm{x}_i'\beta + \delta z_i)]^{y_i} [1 - F(\mathrm{x}_i'\beta + \delta z_i)]^{1-y_i} g(z_i) dz_i \tag{2}$$

We assume that the prior distribution of the coefficients is normal, i.e., $\beta_j \sim N(0,10^{10})$, $\forall_j = 1, \ldots, k$, and $\delta \sim N(0,10^{10})$. These noninformative prior distributions with a very large variance reflect the absence of prior knowledge about the parameters of interest, and they facilitate comparison with classical models.

Combining this prior structure and the likelihood in Eq. 2, we obtained the posterior distribution of parameters (β, δ):

$$\begin{aligned} p(\beta, \delta|\mathrm{y}, \mathrm{x}) &\propto l(\mathrm{y}|\mathrm{x}, \beta, \delta)\pi(\beta, \delta) \\ &= \left\{ \prod_{i=1}^{n} \int_0^{\infty} [F(\mathrm{x}_i'\beta + \delta z_i)]^{y_i} [1 - F(\mathrm{x}_i'\beta + \delta z_i)]^{1-y_i} g(z_i) dz_i \right\} \pi(\beta, \delta) \end{aligned} \tag{3}$$

where $\pi(\beta, \delta)$ is the prior distribution of (β, δ).

We can sample (β, δ) from this posterior distribution by using the WinBUGS package (Windows Bayesian inference Using Gibbs Sampling, developed jointly by the MRC Biostatistics Unit [University of Cambridge, Cambridge, UK] and the Imperial College School of Medicine at St. Mary's, London) [22], based on the Gibbs sampling applying Markov Chain Monte Carlo (MCMC) methods (see Carlin and Polson [23] and Gilks et al. [24] for further details).

One aim of our study is to use logistic regressions in order to make predictions. In Bayesian theory, predictions of future observables are based on predictive distribution. The predictive distribution of unobservable data y_p, given a new set of covariates $\mathrm{x_p} = (x_{p1}, \ldots, x_{pk})$ is defined as

$$p(y_p|\mathrm{y}, \mathrm{x}, \mathrm{x_p}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} l(y_p|\mathrm{x_p}, \beta, \delta) p(\beta, \delta|\mathrm{y}, \mathrm{x}) d\beta d\delta \tag{4}$$

The predictive distribution can also be simulated using MCMC techniques with WinBUGS [22]. We include the WinBUGS code for more details in the Supporting Information Appendix for this article.

Regression Model for Determining the Extension of Hospital Stay due to NI

Frequentist estimation of Poisson-Gamma model. We denote by los_i (length of stay) the number of days that the ith individual remains hospitalized. We then denote by $\mathrm{x_i} = (x_{i1}, \ldots, x_{ik})'$ the vector of factors for the ith individual. Finally, we denote by x_i^{NI} a variable indicating the presence of infection in the ith individual; this variable takes the value one if NI is present, and zero, if otherwise.

We consider a Poisson-Gamma model in which $los_i \sim Poisson(\nu_i\mu_i)$, so

$$P[los_i = y|\nu_i, \mu_i] = e^{-\nu_i\mu_i} \frac{(\nu_i\mu_i)^y}{y!}, \; y = 0, 1, 2, \ldots$$

where

$$\mu_i = exp(\mathrm{x}_i'\beta + \beta^{NI} x_i^{NI}) \tag{5}$$

and ν_i is a parameter of the model that represents a factor of individual heterogeneity, with an individual value for each patient. Values of ν far from 1 indicate that the ith patient presents individual characteristics that explain the length of hospital stay and that are not included in the model. The vector β and the parameter β^{NI} are the coefficients of the covariates $\mathrm{x_i}$ and the indicator variable of infection x_i^{NI}, respectively.

NI is featured among the risk factors related because if the only difference between two ith and i'th individuals is the presence of infection in the first of these, the ratio between the average hospital stay of the two after entry is given by $exp(\beta^{NI})$. Therefore, when the parameter β^{NI} is known, it is possible to estimate the ratio between the average hospital stay of two individuals who are identical except that one of them has NI. Furthermore, this expression represents the pure hazard, i.e., given the covariates, the differences between the values of los_i for individuals with the same values on covariates are random.

To introduce the possibility of heterogeneity not explained by factors in the model, it is considered that ν is a random variable with distribution $Gamma(\alpha, \alpha)$, with density

$$p(\nu|\alpha) = \frac{\alpha^{\alpha} \nu^{\alpha-1} e^{-\alpha\nu}}{\Gamma(\alpha)}$$

By specifying a gamma distribution for ν with shape and scale parameters to be equal, the Negative Binomial (NB) model is derived [25]. As in the classical NB model, ν follows a gamma distribution with $E[\nu] = 1$ and $Var[\nu] = 1/\alpha$.

Thus, $los_i \sim NB(\alpha, \mu)$, i.e.,

$$P[los_i = y|\alpha, \mu_i] = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha)} \left(\frac{\alpha}{\alpha + \mu_i} \right)^{\alpha} \left(\frac{\mu_i}{\alpha + \mu_i} \right)^{y} \tag{6}$$

The model estimation is performed by optimizing the likelihood function using numerical methods, and specifically, R software (written by Robert Gentleman and Ross Ihaka, Statistics Department, University of Auckland, New Zealand) and the MASS package (Modern Applied Statistics in S package devel-

oped by W. N. Venables [CMIS Environmetrics Project, Australia] and B. D. Ripley [Department of Statistics, University of Oxford, Oxford, UK]) [26].

Bayesian estimation of Poisson-Gamma model. As an alternative to the frequentist point of view of the Poisson-Gamma model, we also propose the Bayesian version [27]. Accordingly, we consider $los_i \sim Poisson(v_i\mu_i)$, with $v_i \sim Gamma(\alpha, \alpha)$ and μ_i expressed as in Eq. 5.

Note that in this model v_i again represents individual heterogeneity (not included in the covariates) of each patient, but now we can estimate an individual average value of the heterogeneity for each patient, while in the frequentist estimation only a general average value of the heterogeneity can be estimated.

The hyperparameters β^{NI} and β_j, $\forall j = 1, \ldots, k$, follow prior noninformative normal distribution $N(0,10^{10})$. We propose a flexible hierarchical prior structure for α, $\alpha \sim exp(b)$ and $b \sim exp(0.005)$, where the hyperparameter b follows an exponential distribution with a large variance ($Var(b) = 40000$).

The posterior distribution is obtained combining this prior structure and the likelihood defined in Eq. 6:

$$p(\beta, \beta^{NI}, \alpha, b|\text{los}, \text{x}, \text{x}^{NI}) \propto \left\{\prod_{i=1}^{n} P[los_i|\alpha, \text{x}, \text{x}^{NI}, \beta, \beta^{NI}]\right\}\pi(\beta, \beta^{NI}, \alpha, b) \quad (7)$$

where $\text{los} = (los_1, \ldots, los_n)'$, $\text{x} = (\text{x}_1, \ldots, \text{x}_n)'$, $\text{x}^{NI} = (x_1^{NI}, \ldots, x_n^{NI})'$, and $\pi(\beta, \beta^{NI}, \alpha, b)$ is the joint prior distribution that can be decomposed in

$$\pi(\beta, \beta^{NI}, \alpha, b) = \pi(\beta)\pi(\beta^{NI})\pi(\alpha|b)\pi(b)$$

Again, posterior distributions are obtained by applying MCMC methods implemented by R software, the BRUGS package and WinBUGS software [22]. Source codes are provided in the Supporting Information Appendix for this article at http://www.ispor.org/Publications/value/ViHsupplementary/ViH13i4_Negrin.asp.

4. Results

Predictive Models for NI

The statistical methods consisted of two steps: 1) estimation of logit models and analysis of goodness of fit using information criterion—Akaike information criterion (AIC) for frequentist estimation and deviance information criterion (DIC) for Bayesian estimation and 2) assessment of its predictive accuracy in a split-sample validation cohort [28]. The final sample size was 1021 patients after the elimination of 18 patients with missing values. The entire cohort (1021 patients) was randomly divided into two subcohorts of 766 (75%) and 255 patients (25%). The subcohort of 766 patients was used to develop the logit models. Subsequently, the logit models were externally validated using the remaining 255 patients, who represented the split-sample cohort. The percentage of correct classification, the *c* statistic, and the receiving operating characteristic (ROC) curves are used to quantify the predictive accuracy.

Frequentist estimation of logistic regression models. Three alternative models with different numbers of covariates were fitted (Table 4). This table includes the parameter estimates, the standard errors and *P*-values. A model summary with the sample size,

Table 4 Frequentist estimation of logistic models

Model	Complete			Stepwise			Reduced		
Variable	$\hat{\beta}$	SE	*P*	$\hat{\beta}$	SE	*P*	$\hat{\beta}$	SE	*P*
(Intercept)	1.97	2.90	0.50	2.62	2.03	0.20	−1.63	1.04	0.12
Age	0.02	0.02	0.32						
Sex	−0.25	0.62	0.69						
Length of stay	0.68	0.10	0.00	0.63	0.09	0.00	0.48	0.06	0.00
Admission	0.33	1.14	0.77						
Surgery type	−0.36	1.19	0.76						
Length of surgery	−0.01	0.01	0.23	−0.01	0.01	0.10			
Clean-contaminated	0.06	1.13	0.96	−0.42	0.87	0.63			
Contaminated	−1.59	1.35	0.24	−1.90	1.06	0.07			
Dirty-contaminated	−1.68	1.45	0.25	−2.48	1.12	0.03			
Prophylaxis	−1.06	1.03	0.30						
Preoperative stay	−0.89	0.18	0.00	−0.82	0.14	0.00	−0.69	0.11	0.00
Central tract	−1.29	0.95	0.18	−1.35	0.80	0.09			
Vesical probe	0.39	0.82	0.63						
Nasogastric probe	1.29	0.81	0.11	1.12	0.66	0.09			
Open drainage	−0.58	0.85	0.49						
Artificial respiration	−0.47	1.51	0.76						
Immunosuppressive therapy	2.45	1.12	0.03	1.98	0.96	0.04	2.28	0.79	0.00
Coma	−0.25	1.73	0.89						
Neoplasy	−0.29	0.87	0.74						
COPD	0.21	0.85	0.81						
Chronic hepatopathy	−0.93	1.11	0.40						
Hypoproteinemia	4.00	1.20	0.00	3.85	1.06	0.00	2.05	0.80	0.01
Obesity	0.83	0.68	0.22						
Infection at admission	2.61	0.99	0.01	2.77	0.90	0.00	0.89	0.48	0.06
Number of diagnoses	−0.58	0.35	0.10	−0.70	0.32	0.03			
Model summary									
n		766			766			766	
AIC		168.12			147.48			155.79	
% correct predictions		96.08			96.08			97.25	
c statistic		0.987			0.986			0.987	

Parameter estimates, SE, and *P*-values (*P*).
AIC, Akaike information criterion; COPD, chronic obstructive pulmonary disease; SE, standard errors.

Table 5 Bayesian estimation of symmetric and asymmetric full logistic models

	Symmetric			Asymmetric		
Variable	Mean	SD	CI (95%)	Mean	SD	CI (95%)
Intercept	−7.01	1.41	(−9.94, −4.41)	−27.52	9.98	(−46.50, −7.86)
Delta	—	—	—	−64.04	5.20	(−69.82, −50.63)
Age	−0.00	0.02	(−0.04, 0.03)	−0.27	0.17	(−0.63, 0.05)
Sex	1.15	0.60	(−0.00, 2.37)	4.10	5.08	(−6.26, 13.32)
Length of stay	0.69	0.09	(0.53, 0.88)	7.57	1.03	(5.58, 9.68)
Admission	−0.76	1.12	(−2.98, 1.42)	1.02	7.10	(−12.49, 13.75)
Surgery type	0.62	1.17	(−1.64, 2.92)	3.83	6.99	(−11.04, 14.38)
Length of surgery	−0.00	0.01	(−0.01, 0.01)	−0.02	0.06	(−0.14, 0.10)
Clean-contaminated	0.91	1.06	(−1.12, 3.05)	0.47	6.49	(−12.16, 12.65)
Contaminated	−1.51	1.26	(−4.00, 0.94)	−8.06	5.41	(−14.74, 5.03)
Dirty-contaminated	−0.92	1.26	(−3.43, 1.55)	−4.35	6.39	(−14.31, 9.44)
Prophylaxis	−1.46	0.99	(−3.37, 0.52)	−6.60	5.67	(−14.56, 6.42)
Preoperative stay	−0.93	0.16	(−1.25, −0.63)	−9.18	1.50	(−12.18, −6.33)
Central tract	−1.01	0.92	(−2.83, 0.77)	−6.48	6.17	(−14.66, 8.02)
Vesical probe	−0.61	0.81	(−2.22, 0.95)	−4.23	6.28	(−14.17, 9.24)
Nasogastric probe	1.30	0.78	(−0.22, 2.84)	3.81	6.24	(−9.50, 13.98)
Open drainage	−0.59	0.80	(−2.18, 0.96)	−6.42	5.34	(−14.47, 5.23)
Artificial respiration	−0.75	1.56	(−3.80, 2.37)	−1.50	8.07	(−14.32, 13.63)
Immunosuppressive therapy	2.95	1.14	(0.75, 5.20)	8.70	5.15	(−4.12, 14.79)
Coma	0.65	1.71	(−2.81, 3.90)	−2.18	8.02	(−14.34, 13.38)
Neoplasy	−0.09	0.80	(−1.68, 1.45)	−5.04	6.13	(−14.39, 8.30)
COPD	−0.21	0.84	(−1.91, 1.39)	3.09	6.27	(−9.94, 13.90)
Chronic hepatopathy	−0.49	1.13	(−2.78, 1.64)	0.76	7.55	(−13.34, 13.86)
Immunodeficiency	−2.93	2.09	(−7.27, 0.96)	−2.06	8.32	(−14.42, 13.75)
Hypoproteinemia	3.91	1.18	(1.58, 6.24)	22.01	6.31	(6.42, 29.72)
Obesity	1.23	0.63	(0.01, 2.48)	11.40	5.64	(0.70, 22.84)
Infection at admission	2.67	0.95	(0.88, 4.60)	8.67	5.01	(−3.65, 14.78)
Number of diagnoses	−0.84	0.35	(−1.56, −0.18)	−2.45	2.85	(−7.98, 3.27)
Model summary						
n		766			766	
DIC		205.00			198.09	
% correct predictions		96.47			97.25	
c statistics		0.987			0.990	

Posterior means, SD, and 95% CI.
CI, credibility intervals; COPD, chronic obstructive pulmonary disease; DIC, deviance information criterion; SD, standard deviations.

AIC, percentage of correct classifications, and *c* statistics is also provided. It should be noted that 5 of the 30 covariates considered in this study (peripheral tract, closed drainage, kidney failure, diabetes, and immunodeficiency) could not be included in the models due to the limited number of infected cases and/or problems of colineality.

The first model is the full model, including all the covariates. The AIC is 168.12. This model provides a rate of correct classification of 96.08% (10 errors, 7 false positives, and 3 false negatives). The factors length of stay, preoperative stay, immunosuppressive therapy, hypoproteinemia, and infection at admission are found relevant at the 5% significance level.

The second model, which we term the stepwise model, was obtained by the application of a backward variable selection method, in order to reduce the number of covariates. The resultant model includes 12 of the 25 variables contained in the full model, namely: the length of hospital stay, the duration of surgery, the three covariates corresponding to the degree of contamination, the preoperative stay, central tract, nasogastric probe, immunosuppressive therapy, hypoproteinemia, infection at admission, and the number of diagnoses. The AIC for this model is 147.48. This model, too, provides 96.08% of correct classification for the split-sample (10 errors, 6 false positives, and 4 false negatives).

Finally, we considered a third model, including only those covariates that turned out to be significant (*P*-value smaller than 5%) in the full model. The AIC, in this case, is 155.79. As we can see, the stepwise model has the lower AIC, indicating the best fitting. For the three models considered, the reduced model, with only five covariates, provides the best correct prediction rate (97.25%, seven errors, four false positives, and three false negatives).

Prediction accuracy is also measured by the area under the ROC curve, also known as *c* statistic. The three models estimated show a very similar value for this statistic (0.987 for the complete model, 0.986 for the stepwise model, and 0.987 for the reduced model).

Bayesian estimation of logistic regression models. Bayesian estimation of the logit models involved two steps: firstly, we estimated the full model, for both the symmetric and the asymmetric links. Table 5 shows the results. Posterior mean, standard deviation, and 95% credibility intervals are provided. Then, after having verified the advantages of the asymmetric model, we fitted two reduced asymmetric models (Table 6): the first one included the covariates that were found to be relevant predictors of NI in the symmetric Bayesian logit (seven covariates); the second one included the relevant covariates for the asymmetric Bayesian logit (four covariates). We considered a variable to be relevant as a predictor of NI when the zero value is not included in the 95% credibility interval. The posterior distribution was simulated using WinBUGS [22]. A total of 100,000 iterations were carried out (after a burn-in period of 100,000 simulations). Three different chains were carried out and the convergence was evaluated for all parameters using several tests provided within the WinBUGS Convergence Diagnostics and Output Analysis software.

The frequentist and Bayesian estimations of the complete symmetric model coincide in defining some covariates as relevant

Table 6 Bayesian estimation of reduced asymmetric logistic models

	Reduced 1			Reduced 2		
Variable	Mean	SD	CI (95%)	Mean	SD	CI (95%)
Intercept	−31.60	8.20	(−47.10, −15.94)	−34.77	7.55	(−48.17, −19.54)
Delta	−69.70	8.57	(−79.68, −48.43)	−66.46	10.61	(−79.49, −40.47)
Length of stay	6.69	1.19	(4.27, 8.99)	5.67	1.07	(3.39, 7.52)
Preoperative stay	−9.48	1.90	(−13.23, −5.79)	−7.78	1.62	(−10.82, −4.48)
Immunosuppressive therapy	19.00	10.09	(−2.96, 34.13)	—	—	—
Hypoproteinemia	18.35	9.88	(−2.71, 33.87)	18.39	9.25	(−0.19, 33.74)
Obesity	9.26	4.77	(−0.06, 18.29)	9.87	4.41	(1.34, 18.31)
Infection at admission	8.29	5.30	(−2.11, 18.27)	—	—	—
Number of diagnoses	−5.94	2.70	(−11.39, −0.080)	—	—	—
Model summary						
n		766			766	
DIC		69.07			53.54	
% correct predictions		100			96.47	
c statistic		1			0.990	

Posterior means, SD, and 95% CI.
CI, credibility intervals; DIC, deviance information criterion; SD, standard deviations.

predictors of NI (length of hospital stay, preoperative stay, immunosuppressive therapy, hypoproteinemia, and infection at admission), although some differences can be found in the estimates of the parameters. Furthermore, under the Bayesian approach model, obesity and the number of diagnoses are also considered relevant covariates. In general, the standard errors in the Bayesian models are slightly smaller. As a goodness-of-fit measure, we make use of DIC [29]. The DIC for the complete symmetric model is 205.00, although this criterion is not comparable to AIC. The percentage of correct predictions is slightly larger for the Bayesian approach (96.47%, nine errors, six false positives, and three false negatives).

With the Bayesian estimation of the full asymmetric model, the coefficient of asymmetry δ is both relevant and negative. This coefficient increases the probability of the patient not suffering infection—the largest group. The statistical relevance of this coefficient highlights the importance of considering the asymmetry in the logit model. There are important differences in the estimates of the coefficients with respect to those obtained with the symmetric model. The asymmetric model reduces the number

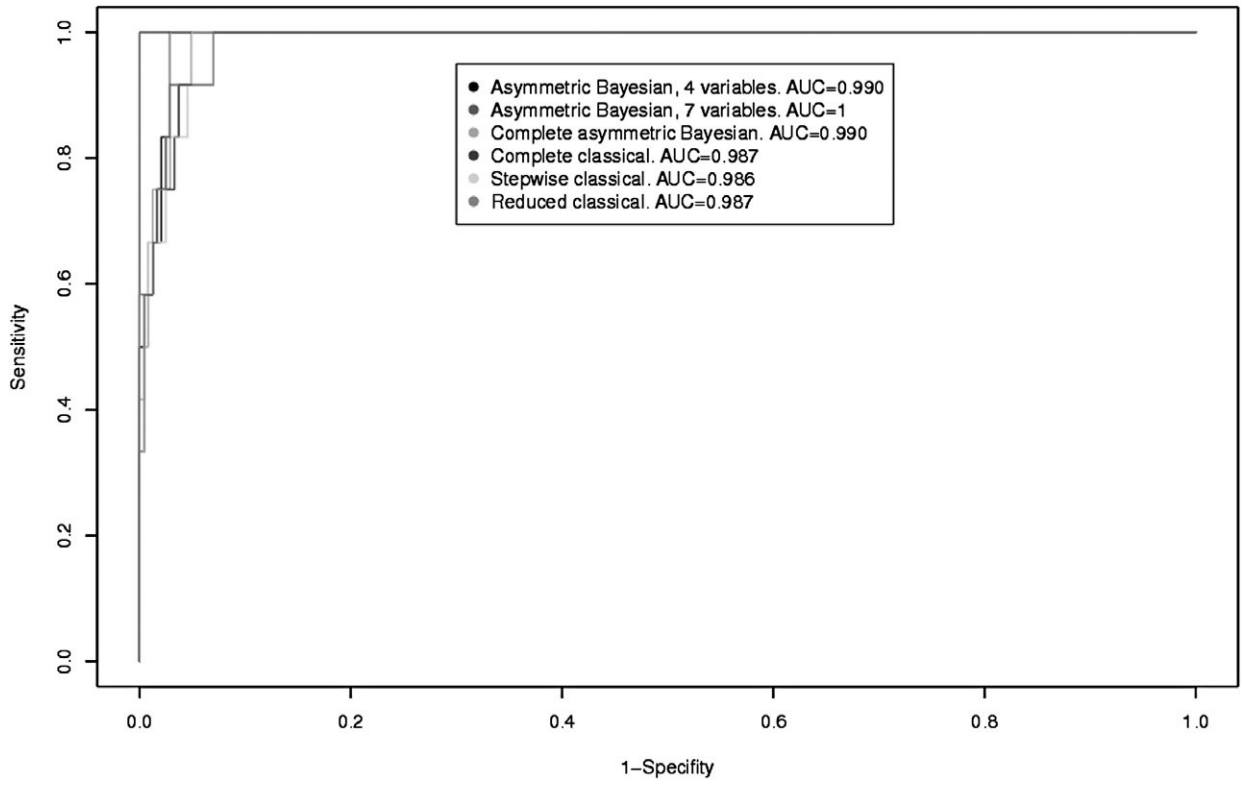

Figure 1 ROC curves and c statistics for proposed models.

of relevant covariates, eliminating immunosuppressive therapy, infection at admission, and number of diagnoses. Using the DIC criterion, the asymmetric model is preferred to the symmetric one, with a DIC of 198.09 versus 205.00 for the symmetric model. The asymmetric Bayesian logit model also presents a higher percentage of correct classifications of infected patients, 97.25% (seven errors, four false positives, and three false negatives) and a higher *c* statistic (0.99 vs. 0.987).

In the second stage of this study, we estimated two abbreviated asymmetric models (Table 6). Both models present very important coefficients of asymmetry δ. The DIC for both these models are clearly smaller than for the full model: 69.07 for the first model with seven covariates (Reduced) and 53.54 for the second one with only four covariates (Reduced 2). It is important to emphasize that the model with the seven covariates found to be significant in the full symmetric model provides 100% of the correct classification. The second abbreviated model, with only four covariates (those significant in the full asymmetric model), still provides 96.47% of the correct classification (nine classification errors, five false positives, and four false negatives).

Figure 1 shows the ROC curves and the *c* statistics according to six models: three frequentist logistic regressions and three asymmetric Bayesian logistic regressions, the full one and the two abbreviated models. The cutoff point to predict a patient with NI is fixed at 0.5. It is important to point out that all the asymmetric models have a predictive capacity better than the best of the frequentist estimations.

Variations in Length of Hospital Stay due to NI

Frequentist estimation of Poisson-Gamma model. As a first approach to the problem of the relationship between length of hospital stay and NI, we estimate a full Poisson-Gamma model for the *los* variable that includes the 30 variables described in section 2. The final sample was 1013 patients after the elimination of 26 cases with missing values. This model was fitted using the maximum likelihood method. Subsequently, an abbreviated model was examined, considering only the covariates that were found to be statistically significant at the 5% level. The results obtained for both models are shown in Table 7.

Based on the AIC criterion, the abbreviated model, with only 14 covariates, is preferred to the full model. The AIC for the full model is estimated to be 3893.02, whereas that for the abbreviated model is 3883.71.

In this section, we are interested in the effect of NI on the duration of hospital stay. For both models, the variable *NI* is statistically significant. In the abbreviated model, the coefficient of *NI* is estimated to be 1.03 units. This means that the average length of hospital stay for a patient with an infection will be multiplied by a factor of $e^{1.03} = 2.80$ in comparison to a noninfected patient with the same characteristics.

Bayesian estimation of Poisson-Gamma model. The analysis was complemented with the Bayesian estimation of the Poisson-Gamma models. As in the previous Bayesian estimation of logit models, MCMC techniques were used to estimate the posterior distributions of the parameters of interest. Three chains of 100,000 samples were recorded after a burn-in sample of 100,000. Different diagnoses were carried out to ensure the desired convergence of the simulations.

Table 8 shows the results of the Bayesian estimation of the full Poisson-Gamma model and that of the abbreviated model, which includes only the relevant covariates. The goodness of fit for both Bayesian models was analyzed using the DIC. The full model is preferred to the abbreviated one with a value of DIC of 3534.11 versus 3537.65 for the abbreviated model. The posterior mean for the coefficients of the relevant covariates are similar in both models.

Table 7 Frequentist estimation of full and reduced Poisson-Gamma regression models for length of stay data

Model	Complete			Reduced		
Poisson model	$\hat{\beta}$	SE	*P*	$\hat{\beta}$	SE	*P*
(Intercept)	−0.46	0.54	0.39	−0.27	0.10	0.01
NI	1.04	0.10	0.00	1.03	0.10	0.00
Age*	0.17	0.04	0.00	0.18	0.04	0.00
Sex	0.12	0.07	0.07			
Admission	0.30	0.13	0.02	0.28	0.13	0.03
Surgery type	0.55	0.14	0.00	0.61	0.13	0.00
Length of surgery*	0.26	0.04	0.00	0.28	0.04	0.00
Clean-contaminated	0.15	0.11	0.16			
Contaminated	−0.06	0.13	0.61			
Dirty-contaminated	−0.33	0.13	0.01	−0.36	0.08	0.00
Prophylaxis	0.23	0.10	0.02	0.25	0.09	0.01
Preoperative stay*	0.29	0.03	0.00	0.29	0.03	0.00
Peripheral tract	0.09	0.53	0.86			
Central tract	0.08	0.13	0.55			
Vesical probe	0.14	0.10	0.17			
Nasogastric probe	0.40	0.10	0.00	0.48	0.09	0.00
Open drainage	0.43	0.09	0.00	0.44	0.08	0.00
Closed drainage	0.48	0.10	0.00	0.46	0.09	0.00
Artificial respiration	−0.16	0.24	0.52			
Immunosuppressive therapy	−0.26	0.19	0.18			
Coma	−0.14	0.25	0.57			
Kidney failure	−0.05	0.26	0.85			
Diabetes	0.22	0.10	0.03	0.27	0.10	0.01
Neoplasy	0.31	0.11	0.00	0.27	0.11	0.01
COPD	0.03	0.10	0.78			
Chronic hepatopathy	0.10	0.14	0.45			
Immunodeficiency	−0.04	0.33	0.91			
Hypoproteinemia	−0.30	0.15	0.05			
Obesity	0.11	0.08	0.19			
Infection at admission	0.07	0.11	0.52			
Number of diagnoses	0.09	0.04	0.04	0.09	0.04	0.02
Gamma model	$\hat{\alpha}$	SE	*P*	$\hat{\alpha}$	SE	*P*
α	2.72	0.26	0.00	2.59	0.24	0.00
Model summary						
n		1013			1013	
AIC		3893.02			3883.71	

*These variables have been standardized.
Parameter estimates, SE, and *P*-values (*P*).
AIC, Akaike information criterion; COPD, chronic obstructive pulmonary disease; NI, nosocomial infection; SE, standard errors.

The results obtained by the Bayesian estimations in this section are similar to those obtained with the frequentist ones. In particular, the coefficients for the existence of NI are statistically significant both in the full and the abbreviated Bayesian models. The estimation of the posterior mean for the β^{NI} coefficient in the full model is 1.05, versus 1.04 for the abbreviated model. To interpret these coefficients, we need to calculate their exponential transformation, from which we conclude that the existence of NI would multiply the length of hospital stay by a factor of $e^{1.05} = 2.87$ and by $e^{1.04} = 2.83$ for the full and abbreviated models, respectively.

In addition, with the Bayesian approach it is possible to specify an individual distribution for the parameter v_i that refers to the heterogeneity of the sample. Using the results from the abbreviated model, we found that only 79 of the 1013 (7.8%) individuals of the sample showed a 95% Bayesian interval for v_i that excludes the value 1, indicating significant individual heterogeneity.

5. Discussion

We have proposed the use of the Bayesian approach of logit models with an asymmetric link to estimate the NI probability of

Table 8 Bayesian estimation of full and reduced Poisson-Gamma regression models for length of stay data

	Complete model			Reduced model		
Poisson model	Mean	SD	CI (95%)	Mean	SD	CI (95%)
Intercept	−1.31	0.55	(−2.43, −0.10)	−0.28	0.10	(−0.47, −0.08)
NI	1.05	0.11	(0.84, 1.27)	1.04	0.10	(0.84, 1.25)
Age*	0.17	0.04	(0.10, 0.24)	0.18	0.03	(0.11, 0.25)
Sex	0.12	0.07	(−0.01, 0.25)			
Admission	0.31	0.14	(0.04, 0.59)	0.27	0.13	(0.02, 0.54)
Surgery type	0.55	0.14	(0.26, 0.83)	0.61	0.14	(0.34, 0.88)
Length of surgery*	0.27	0.05	(0.19, 0.36)	0.28	0.04	(0.20, 0.36)
Clean-contaminated	0.15	0.11	(−0.06, 0.36)			
Contaminated	−0.07	0.13	(−0.32, 0.18)			
Dirty-contaminated	−0.34	0.13	(−0.59, −0.09)	−0.36	0.08	(−0.53, −0.20)
Prophylaxis	0.23	0.10	(0.05, 0.42)	0.25	0.09	(0.06, 0.43)
Preoperative stay*	0.29	0.04	(0.21, 0.37)	0.29	0.04	(0.22, 0.37)
Peripheral tract	0.02	0.54	(−1.12, 1.15)			
Central tract	0.07	0.13	(−0.18, 0.32)			
Vesical probe	0.14	0.10	(−0.05, 0.34)			
Nasogastric probe	0.40	0.10	(0.20, 0.60)	0.47	0.09	(0.29, 0.65)
Open drainage	0.43	0.09	(0.26, 0.60)	0.44	0.08	(0.28, 0.60)
Closed drainage	0.48	0.10	(0.29, 0.68)	0.46	0.10	(0.27, 0.65)
Artificial respiration	−0.15	0.26	(−0.67, 0.37)			
Immunosuppressive therapy	−0.25	0.20	(−0.65, 0.15)			
Coma	−0.14	0.27	(−0.67, 0.38)			
Kidney failure	−0.04	0.27	(−0.56, 0.51)			
Diabetes	0.23	0.11	(0.02, 0.44)	0.27	0.10	(0.07, 0.48)
Neoplasy	0.31	0.11	(0.08, 0.53)	0.27	0.11	(0.06, 0.48)
COPD	0.03	0.10	(−0.17, 0.22)			
Chronic hepatopathy	0.11	0.15	(−0.18, 0.40)			
Immunodeficiency	−0.03	0.35	(−0.69, 0.67)			
Hypoproteinemia	−0.30	0.16	(−0.62, 0.02)			
Obesity	0.11	0.09	(−0.05, 0.28)			
Infection at admission	0.07	0.12	(−0.16, 0.30)			
Number of diagnoses	0.09	0.04	(0.01, 0.18)	0.09	0.04	(0.01, 0.16)
Gamma model	Mean	SD	CI (95%)	Mean	SD	CI (95%)
α	2.50	0.25	(2.07, 3.03)	2.47	0.24	(2.04, 2.97)
Model summary						
n		1013			1013	
DIC		3534.11			3537.65	

*These variables have been standardized.
Posterior means, SD, and 95% CI.
CI, credibility intervals; COPD, chronic obstructive pulmonary disease; DIC, deviance information criterion; NI, nosocomial infection; SD, standard deviations.

a patient undergoing hospital surgery, comparing the reliability of these estimates with that provided by the frequentist version of logistic regression models.

It should be emphasized that the Bayesian methodology establishes clear differences, even between the symmetric logit model and its analog in the classical methodology, fitted by the maximum likelihood method. These differences are observed not only in obtaining estimates, but also in the significant variables established in the two models. For instance, obesity and the number of diagnoses are not relevant factors in the classical analysis but they are so in the Bayesian analysis. Nevertheless, the estimates are similar for the common significant variables, although the standard errors are, in general, slightly lower in the Bayesian estimation of the logistic regression model.

Comparing the logit model (with a symmetric link) and the skewed logit model, we observe clear differences in the detection of significant variables: seven variables are significant in the first model, versus only four in the second; since immunosuppressive therapy, infection at admission and number of diagnoses are eliminated. Nevertheless, the great advantage of these skewed logit models is their great capacity for discrimination (as can be seen in Fig. 1), correctly classifying 100% of patients with NI. This discrimination capacity seems to show that the asymmetry node makes it possible to obtain a more accurate fit for data with different proportions of zeros and ones.

In addition to logistic regression, there are several other approaches to the problem of how to formally model the relationship between the probability of an event and a set of covariates, such as a probit analysis. Furthermore, an important variant of this class of problems arises when interest is not only in whether the event of interest occurs or not, but also in the time until the event occurs. The body of methods for analyzing such data is known as survival analysis [30].

Secondly, we proposed the use of Poisson-Gamma regression models as a multivariate procedure for identifying factors that really are related to a lengthening of hospital stay. Case-control studies are usually employed to estimate differences between infected and noninfected patients. Propensity Score Matching can be used to create groups of treated and control units that have similar characteristics and so comparisons can be made within these matched groups [31]. Nevertheless, regression models do allow us to distinguish the variables that really, and in a multivariate way, influence the lengthening of hospital stay. Likewise, they make it possible to evaluate the relative differences between the average hospital stay of infected and noninfected patients with the same conditions as for the other variables. In fitting these Poisson-Gamma regression models, we considered both the maximum likelihood method and the Bayesian techniques. It should be noted that, unlike the logit models, there are hardly any differences between the fits. Nevertheless, the Bayesian model has the advantage of providing a random model for the heterogeneity of each individual, which allows us to analyze the characteristics of the most atypical cases in the model.

Authors thank the editor and three anonymous referees for constructive comments and suggestions.

Source of financial support: This research has been partially support by the grant SEJ2006-12685 (Ministerio de Educación y Ciencia (MEC), Spain).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix. WinBUGS codes.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

1 Plowman RP, Graves N, Roberts JA. Hospital Acquired Infection. London: Office of Health Economics, 1997.
2 EPINE Working Group. Prevalence of hospital-acquired infections in Spain. J Hosp Infect 1992;20:1–13.
3 Vaque J, Rosell J, Trilla A. Nosocomial infections in Spain: results of five nationwide serial prevalence surveys (EPINE project, 1990 to 1994). Nosocomial infections prevalence study in Spain. Infect Control Hosp Epidemiol 1996;17:293–7.
4 EPINE Working Group. Prevalence of hospital infections in Spain. EPINE Study. Spanish Society in Preventive Medicine and Public Health, Madrid, 2008 [in Spanish].
5 Haley R. Incidence and nature of endemic and epidemic nosocomial infections. In: Bennett JV, Brachman P, eds. Hospital Infections. Boston, MA: Little Brown, 1995.
6 Plowman RP, Graves N, Griffin MAS, et al. The rate and cost of hospital-acquired infections occurring in patients admitted to selected specialties of a district general hospital in England and the national burden imposed. J Hosp Infect 2001;47:198–209.
7 Kampf G, Gastmeier P, Wischnewski N, et al. Analysis of risk factors for nosocomial infections—results from the first national prevalence survey in Germany (NIDEP study, part 1). J Hosp Infect 1997;37:103–12.
8 Girou E, Stephan F, Novara A, et al. Risk factors and outcome of nosocomial infections: results of a matched case-control study of ICU patients. Am J Resp Crit Care 1998;157:1151–8.
9 Savas L, Guvel S, Onlen Y, et al. Nosocomial urinary tract infections: micro-organisms, antibiotic sensitivities and risk factors. West Indian Med J 2006;55:188–93.
10 Huang Y, Zhuang S, Du M. Risk factors of nosocomial infection with extended-spectrum beta-lactamase-producing bacteria in a neonatal intensive care unit in China. Infection 2007;35:339–45.
11 Albert H, Chib S. Bayesian analysis of binary and polychotomous response data. J Am Stat Assoc 1993;88:669–79.
12 McCulloch R, Polson N, Rossi P. A Bayesian analysis of the multinomial probit model with fully identified parameters. J Econ 1999;99:173–93.
13 O'Hagan A, Woodward EG, Moodaley LC. Practical Bayesian analysis of a simple logistic regression: predicting corneal transplants. Stat Med 1990;9:1091–101.
14 Albert J, Chib S. Bayesian residual analysis for binary response regression models. Biometrika 1995;82:747–69.
15 Holmes CC, Held L. Bayesian auxiliary variable models for Binary and Multinomial Regression. Bayesian Anal 2006;1:145–68.
16 Ll B, Pérez JM, Ayuso M, et al. A Bayesian dichotomous model with asymmetric link for fraud in insurance. Insur Math Econ 2008;42:779–86.
17 Chen M, Dey D, Shao Q. A new skewed link model for dichotomous quantal response data. J Am Stat Assoc 1999;94:1172–86.
18 Graves N. Economic and preventing hospital acquired infection. Emerg Infect Dis 2004;10:561–6.
19 Garner JS, Jarvis WR, Emori TG, et al. CDC definitions for nosocomial infections, 1988. Am J Infect Control 1988;16:128–40.
20 Basu S, Mukhopadhayay S. Bayesian analysis of binary regression using symmetric and asymmetric links. Sankhya 2000;3:372–87.
21 Bazán J, Franco M, Bolfarine H. A skew item response model. Bayesian Anal 2006;1:861–92.
22 Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS: a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 2000;10:325–37.
23 Carlin BP, Polson NG. Monte Carlo Bayesian methods for discrete regression models and categorical time series. Bayesian Stat 1992;4:577–86.
24 Gilks WR, Richardson S, Spiegelhalter DJ. Introducing markov chain monte carlo. In: Gilks WR, Richardson S, Spiegelhalter DJ, eds. Markov Chain Monte Carlo in Practice. London: Chapman and Hall, 1995.
25 Cameron AC, Trivedi PK. Regression Analysis of Count Data. Econometric Society Monographs. Cambridge: Cambridge University Press, 1998.
26 R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing, 2008. Available from: http://www.R-project.org [Accessed August 15, 2009].
27 Winkelmann R. Econometric Analysis of Count Data (4th ed.). Berlin: Springer-Verlag, 2003.
28 Steyerberg WE, Harrell FE, Borsboom JM, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001;54:774–81.
29 Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit. J Roy Stat Soc B Met 2002;64:583–639.
30 Klei DG. Survival Analaysis: A Self-Learning Text (Statistics for Biology and Health) (2nd ed.). New York: Springer, 2005.
31 Baser O. Too much ado about propensity score models? Comparing methods of propensity score matching. Value Health 2006;9:377–85.